DOCUMENT RESUME

ED 166 601 CG 013 211

AUTHOR Apking, Theodore D.
TITLE A Behavior and Cost Analysis of Two Remediation Procedures.
PUB DATE 78
NOTE 68p.; Masters Thesis, Western Michigan University

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Cost Effectiveness; Higher Education; *Instructional Improvement; *Instructional Student Costs; Performance Criteria; *Remedial Instruction; *Student Motivation; *Teaching Techniques

ABSTRACT

Two group-comparison experiments indicate that it may be more cost efficient not to include a remediation procedure in a course, even though it may be to the students' advantage to remediate concepts they did not master in an undergraduate university course. In the first experiment, students in the remedial group received minimal course credit for first-attempt quiz scores of less than 90% correct; they then had the opportunity to earn full credit for mastering the unit at 90% correct on the remedial quiz. The students in the control group kept their first-attempt quiz scores and were not able to remediate. The results of this experiment showed no difference in student performance between groups. In the second experiment, participation in the weekly remedial activities was optional for students in the remedial group who scored 70% or above on the first quiz attempt. Students in the control group could not remediate and kept their first quiz score. In this experiment, students in the remedial group scored 34% higher than students in the control group on an individual probe test consisting of quiz questions students missed on the first quiz, but no difference occurred on a comprehensive achievement test. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166601

CG 013211

A BEHAVIOR AND COST ANALYSIS OF TWO REMEDIATION PROCEDURES

Theodore D. Apking, M.A.

Western Michigan University, 1978

Two group-comparison experiments were conducted to evaluate the effectiveness of two remedial quizzing procedures in an undergraduate university course. In Experiment 1, students in the remedial group received minimal course credit for first-attempt quiz scores of less than 90% correct; they then had the opportunity to earn full credit for mastering the unit at 90% correct on the remedial quiz. The students in the control group kept their first attempt quiz score and were not able to remediate. The results of Experiment 1 showed no difference in student performance between groups.

In Experiment 2, students in the remedial group had the option to participate in the weekly remedial activities if they scored 70% or above on the first quiz attempt. In addition, the remedial activities now required active student participation. As before, students in the control group could not remediate and kept their first quiz score. In Experiment 2, students in the remedial group scored 34% higher than students in the control group on an individualized probe test consisting of quiz questions students missed on the first quiz, but no difference occurred on a comprehensive achievement test.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. D. Apking

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CG 013211

ACKNOWLEDGEMENTS

I would like to thank the staff of Psychology 351, especially Doug Nicola, for their continual encouragement and help while collecting data for this research. Dr. Joel Bowman deserves thanks for his critical editing of my many rough drafts; and Dr. Howard Poole for the many hours of guidance and data analysis. Ultimately my most sincere gratitude goes to Dr. Richard Malott who arranged for me to consult with two excellent behavior analysts, Terry McSween and Tim Wysocki.

Theodore David Apking

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
II	GENERAL METHOD	6
	Course Description	6
	Dependent Variables	8
	Test Over Remediation Procedure	11
	Reliability of the Dependent Variables	11
III	EXPERIMENT 1	14
	Method	14
	Results and Discussion	15
IV	EXPERIMENT 2	31
	Method	33
	Results and Discussion	36
V	CONCLUSIONS	57
VI	REFERENCES	58
VII	APPENDICES	60
	Appendix A: Cost Analysis	60
	Appendix B: Recommendations for Psychology 351	62
	Appendix C: Recommendations for Future Research	63

A BEHAVIOR AND COST ANALYSIS OF TWO REMEDIATION PROCEDURES

by

Theodore D. Apking

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment
of the
Degree of Master of Arts

Western Michigan University
Kalamazoo, Michigan
August 1978

INTRODUCTION

Instructors include remediation procedures in their course design on the assumption that students have the opportunity to learn more when remediation is available. Several studies support the assumption of increased learning through remediation. Bostow and Blumenfeld (1972), Bostow and O'Connor (1973) and DuNann and Fernald (1976) all report that students participating in remediation procedures scored significantly higher on final examinations than students not participating in remediation procedures. In addition, students rate courses which include remediation procedures favorably on subjective course evaluations (Malott and Svinicki, 1969).

When evaluating the merit of educational procedures such as remediation, we should examine the student's terminal performance, the cost of implementation, and the reliability of the dependent and independent variables (Hursh, 1976). If students performances are improved reliably and the cost of implementation is judged to be reasonable, the procedure may be cost-efficient (see Appendix A).

Students attained only small increments of improvement in studies of the efficacy of remediation. Bostow and O'Connor show approximately 4% improvement on the final examination as a result of remediation, and DuNann and Fernald report that remediation accounted for approximately 5% to 8% of the total variance in scores on the final examination. When considering the small mean improvement in final examination performance (6% - 15%) in comparisons of Personalized

Systems of Instruction (PSI) and lecture courses (Hursh, 1976), it is not surprising that remediation procedures alone produce even smaller amounts of improvement in students' terminal performance.

Because the improvement in learning is small, the additional cost of implementing a remediation procedure becomes especially critical. Decreasing amounts of improvement are characteristic as student performances approach their peak, making close monitoring of costs necessary. Cost-efficiency estimates prove to be useful as they allow comparing, in terms of cost, the performances produced by two or more procedures.

In view of the small improvement in learning remediation procedures produced, another concern is the reliability of the measurement system. Small shifts in the value of a dependent variable as a result of poor reliability can easily mask a small effect or cause an effect to be erroneously attributed to the manipulation of the independent variable. It is difficult to assess from the current literature what portion of the reported difference in performance between remediation and no-remediation procedures is a result of the remediation procedure as the reliability data are not compelling, or even reported. For example, Hursh (1976) reported in a literature review that approximately two thirds of the published research on PSI failed to report reliability data on performance measures and no data appear on the reliability of the manipulation of the independent variable. Similarly, studies of remediation in lecture courses failed to report adequate reliability procedures (Bostow and Blumenfeld, 1972; Bostow and O'Connor, 1973; Clayton and Madsen, 1974).

A brief review of research follows covering the major findings on remediation procedures in lecture courses. Bostow and Blumenfeld (1972) and Bostow and O'Connor (1973) evaluated a remediation procedure including a point contingency to motivate students who failed to master the material on the first attempt. The motivational-point system specified that students who failed to master the material at 90% correct on the first attempt received minimal course credit and had the option to improve their score on a second form of the quiz. Students who remediated received the sum of the points they earned on the first and second quizzes. If they achieved 70% or more on the first quiz, full credit for the week's work was possible by mastering the unit on the second attempt. The researchers did not include additional learning activities between the first and second quiz attempts. Nevertheless, data from both studies indicated significant differences between final examination scores in favor of the remedial procedure.

DuNann and Fernald (1976) also found remediation superior to no-remediation in a lecture course. Their procedure included a motivational system called a modified "Dooms Day Contingency". Students working under the contingency had to maintain a score of 80% correct on each unit or drop the class.

Clayton and Madsen (1974) attempted to evaluate several components of contingency managed instruction in a lecture course with a "Progressive Multiple Baseline Achievement Test" (Miller, 1972). They failed to replicate the results of studies reporting superior performance on final examinations as a function of a remediation procedure.

Sundberg (1977) manipulated mastery criteria and the opportunity

to retest, in a PSI course, and the results indicated no improvement on review quizzes when students had the opportunity to remediate daily work.

In view of the repeated failure of researchers to produce a difference in the terminal performance of students exposed to remediation, I attempted to systematically replicate (repeat an experiment using the same basic procedure with minor changes; Sidman, 1960) the procedure previously determined effective by Bostow and O'Connor in 1973.

Several features differed between the present study and the Bostow and O'Connor study. They assigned negative course credit to students who scored 60% or less credit on the first quiz. The present study's point system did not include such a feature. Their course design provided optional lectures between the first and second quizzes; alternatively, in the present course design, optional tutoring sessions were available at convenient times throughout the week. In an attempt to make remediation aversive, Bostow and O'Connor gave more difficult forms of the weekly quiz during remediation. The relative difficulty of the remedial and no-remedial quizzes was not systematically manipulated in the present study. Added features of the present study were a quiz assessing the extent to which the students in the remedial group were familiar with guidelines for the remediation procedure and reliability checks on all performance measures.

The experimental design remained a group-comparison, and the major dependent variable was students' scores on a multiple-choice achievement test over a sample of the course material.

All performance measures will be considered output and all staff costs input in a cost-efficiency estimation for each procedure.

GENERAL METHOD

Course Description

Applied Behavior Analysis Lab I is a contingency-managed, instructor-paced course with weekly quizzing over a sequence of reading which begins with previously presented material and gradually moves to more difficult material.

Materials

On the first day of class, teaching apprentices (TAs) distributed the three course manuals which describe the course components. Students were also required to read the following books: Managing Behavior, (R. V. Hall, 1974); How To Draw Graphs, (Katzenberg, 1975); Behavioral Analysis of Everyday Life, (Reese and Woolfenden, 1973); Issues in the Analysis of Behavior, (Malott, General and Snapper, 1973); and Reflex and Operant Conditioning, (Geis, Stebbins and Lundin, 1965).

Lectures

The course instructor lectured over various topics in behavior analysis. On the weekly quiz, one question appeared over lecture content from the previous week's lecture.

Unit Quizzes

The TAs prepared weekly unit quizzes composed of ten short-answer essay questions selected from a pool of questions about topics covered

by the reading objectives. The graduate assistant reviewed each quiz to see that the major issues in the unit had been covered adequately by the questions. Both the first quiz and the remedial quiz underwent this review. In some units, both quizzes contained questions about the same objective or topic, but the same question never appeared on more than one quiz. It was not unusual for a single question to have several parts with each part corresponding to a combination of objectives.

A TA monitored the quiz period and allowed approximately 40 minutes for the students to complete the quiz. Teaching assistants graded the quizzes and posted the scores immediately following the quiz period. Students were able to review the corrected quizzes the next day during the assistants' office hours. Students who found mistakes in grading were able to file a request to have the quiz regraded.

Final Examination

The final examination consisted of 15 short-answer essay questions that previously appeared on the weekly unit quizzes. Students could not receive a course grade higher than the grade they received on the final examination, but they had two opportunities to take the final examination.

Course Evaluation Forms

On the day of the final examination the class completed the standard departmental course evaluation. The form used by the Department of Psychology at Western Michigan University addresses four major areas:

teaching ability of the instructor, course organization and planning, student staff, and materials. All four areas were rated on a five point scale. One question asked students to evaluate the teaching procedures and was used as a measure of the students' satisfaction with their respective quizzing procedure. The students' responses to this question showed whether exposure to the remedial or no-remedial quizzing procedure affected the students' evaluation of the course's teaching procedures.

Teaching Apprentices (TAs)

Throughout the study, I had the aid of five TAs who had been selected on the basis of the excellence of their academic and social skills. The TAs earned academic credit for their weekly duties which included grading quizzes, tutoring students, performing clerical tasks, and writing weekly unit quizzes. Three of the five TAs assisted in a pilot study similar to the present study during the previous semester. Their research duties in the present and past studies consisted of taking reliability checks on grading accuracy and recording data. The other two assistants had served as subjects in the pilot study, and their duties in the present study consisted of grading quizzes.

Dependent Variables

Achievement Tests

The achievement test was used to assess whether a difference in

the students' terminal performance resulted from participation in the experimental remediation procedures. The test consisted of 40 multiple-choice questions which sampled major topics from each weekly unit in the course. These questions had not appeared on any weekly quizzes. In previous semesters, pilot studies showed that the regular course final examination was not an appropriate dependent variable for the present research. Ceiling effects (insensitivity of a dependent variable resulting from little room for improvement within the range of the operant level of responding and the upper limit of the measure) reduced the reliability of the final examination, requiring the construction of the achievement test which produced a wider range of scores around the mean.

Test questions underwent several checks as predictors of overall student achievement of the reading material. One such test involved arranging a list of students in descending order by achievement test score and noting which items high scoring students consistently missed. Then each item missed by more than one-half of the high scorers was rewritten or replaced if judged unclear or misleading. Another test validity, conducted by the Western Michigan University Testing Services Center, included an extensive statistical item analysis by computer in which only items with an index of discrimination (difference between the percent correct of the students in the upper three centiles and the percent correct for the students in the lower three centiles) of between 50 and 80 were retained. A doctoral candidate edited the achievement test for technical and grammatical errors following the formal evaluations of the test. It may be of interest to note that

several of the TAs took the achievement test during the first couple weeks of the semester and scored better than the best students, with one exception, enrolled in the course. However, the achievement test may be measuring a third and correlated variable.

Students in the remedial and no-remedial groups completed the achievement test at times specified by the experimental design for each experiment. Achievement tests were unannounced and no points were contingent upon the level of performance on the test. Students were instructed to "do their best" and received five points toward the next quiz for completing the test.

Weekly Unit Quiz Scores

The TAs recorded the first and second quiz scores as percent correct.

Student Study Time

Throughout the study, students anonymously recorded the amount of time they had studied for the weekly unit quiz. Once a week, the graduate assistant reminded students to record their study time in the appropriate section of the study time form immediately after they had handed in their quizzes.

Staff Time

The TAs recorded the time they spent working on several specified tasks as part of their weekly duties.

Test Over Remediation Procedure

During the first week of remediation, students in the remedial group received a written set of guidelines and procedures for remediation. The students took a brief test over the procedure for remediation after the remediation procedure had been in effect for approximately five weeks.

Reliability of the Dependent Variables

Achievement Test

The graduate assistant used an electronic scorer (Datronics 550/D) to read the students answers from a standard data card marked with a #2 pencil. Then the TAs rescored approximately 10% of the tests by hand to check the accuracy of the electronic scoring machine. No errors in grading occurred.

Weekly Unit Quizzes and Final Examination

Teaching apprentices earned points toward their grades contingent upon meeting a specified percent of agreement with an advanced teaching apprentices (ATAs) when grading as reported by McSween (1977). The ATA directed and monitored the grading-reliability procedure.

The grading-reliability procedure included reviewing the quiz material. During the review, the ATA specified exactly what material was relevant for the quizzes to be graded, and also specified the answer to each quiz question. Further, the ATA indicated on which

questions partial credit might be appropriate and exactly how many points would be reasonable. Occasionally an answer on the key had to be altered by the group to satisfy all of the options for a correct answer. The answer key contained the correct answer and point value for each question.

Following the review, a check for grading reliability occurred--the ATA randomly selected 25% of the quizzes and randomly distributed them to the TAs. Then, the TAs independently graded the sample quizzes without marking on them and recorded the number of points awarded to each question of each quiz on a reliability form. Throughout the reliability check the TAs had access to the answer key as they would when they graded the remainder of the quizzes following the reliability check.

After the TAs finished grading the sample quizzes, the ATA independently regraded each quiz without consulting the TAs' reliability form. Then, the ATA explained to the group all items on which a disagreement occurred on one or more of the reliability forms. Following the explanation, the ATA recorded the number of agreements and disagreements in grading between his grading and the TAs' for each question on the sampled quizzes. An agreement was defined as an instance in which the ATA and the TA awarded exactly the same number of points to a question. A disagreement was defined as an instance in which the ATA and the TA awarded a different number of points to a question. The record of agreements and disagreements in grading served as a basis for the TAs's grades and as a measure of the overall

reliability of grading.

The TAs received different numbers of points for different levels of agreement (grading reliability). When a TA's percent of agreement was 90% or more, the TA received all of the possible 5 points for the day's grading. However, if the TA failed to achieve 90% agreement, he/she received fewer points. The TAs receive three points for 70% to 89% agreement and no points for below 70% agreement. Because the TAs' grades depended on this procedure, the graduate assistant resolved questionable decisions.

The formula used to calculate the exact grading reliability was as follows:

$$\frac{\text{agreements}}{\text{agreements + disagreements}} \times 100 = \text{percentage of agreement}$$

EXPERIMENT 1

Method

Subjects

On the first day of class in Applied Behavior Analysis Lab 1, I asked for volunteers to participate in a study on remediation. The standard procedures did not include remediation. I explained that students who signed an informal consent document and volunteered to participate would be randomly assigned to either the experimental remediation procedure or to the standard no remediation quizzing procedure. Out of a course enrollment of 47, a total of 28 students volunteered. I divided them into two groups with 14 students in each group.

Remedial Group

Students in the remedial group had two opportunities to master the weekly unit and earn a maximum of 10 "grade points" a week. Each quiz contained a total of 30 "raw score points". "Raw score points" were converted into "grade points" for the first quiz as follows:

Raw Score Points	Grade Points
27 - 30	10
21 - 26	2
0 - 20	0

Students who scored 26 "raw score points" or less on the first quiz were able to take the second form of the quiz. The "raw score

points" on the second quiz were converted into "grade points" as follows:

Raw Score Points	Grade Points
27 - 30	8
21 - 26	6
18 - 20	4
0 - 17	0

"Grade points" from the first and second quizzes accumulated to arrive at the weekly total.

No-Remedial Group

Students in the no-remedial group did not have the option of taking the weekly unit quiz a second time. These students retained the number of correct out of a possible 30 points as their score for the unit. All other quizzing conditions remained consistent with the remedial group.

Achievement Test

Students took the achievement test in its entirety on the day of the last weekly quiz during the week before the final examination.

Results and Discussion

The results of Experiment 1 indicate that students significantly improved their weekly unit score when they remediated, and that students in the remedial group earned significantly more points than students in the no-remedial group, on the weekly unit quizzes. No difference in terminal performance occurred.

First Quiz Performance

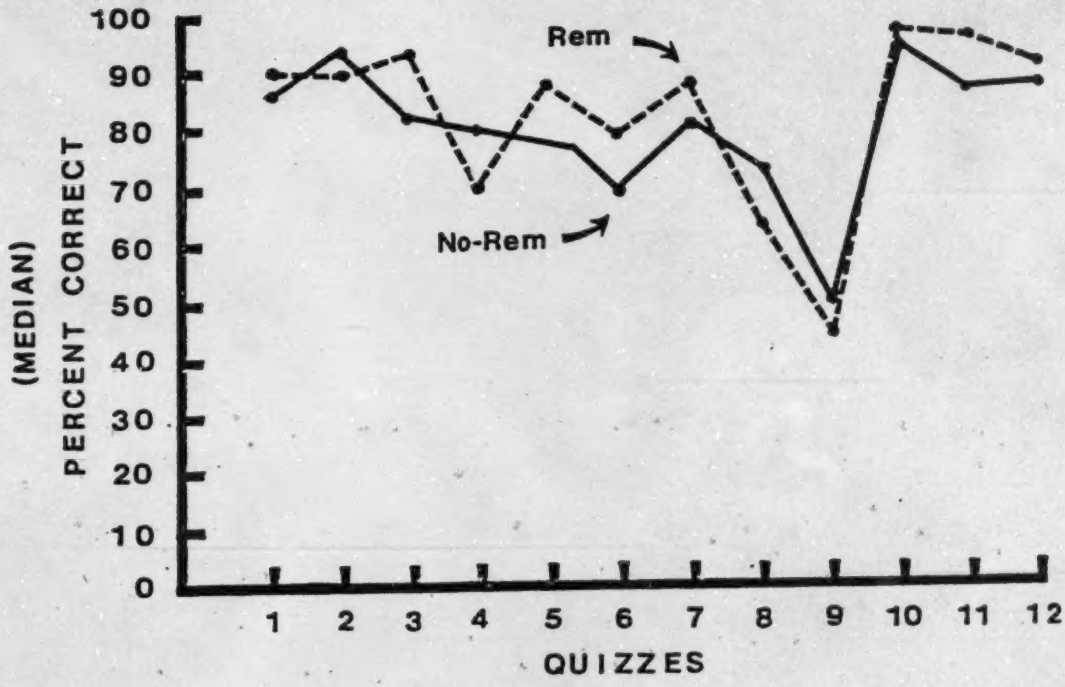
Figure 1 shows the median percent correct on the first weekly unit quizzes for the remedial and the no-remedial groups. Like Clayton and Madsen (1974), Bostow and O'Connor (1973), and Trainor (1977), the data indicate no significant difference between groups across the semester.

Students who participated in the remedial group reported they studied more, on the average, than students in the no-remedial group--two hours and ten minutes compared to one hour and fifty-two minutes for each first quiz.

First quiz performance data suggests that students tended to perform course work at the level necessary to achieve an "A", a finding reported by Sundberg (1977), Johnson and O'Neill (1973), and Semb (1974). The median quiz score approximated the minimum score necessary to maintain a course grade of "A" when averaged with the percent correct from the laboratory section of the course which represented 50% of the course grade before the final examination. It was possible for students to average as little as 80% correct on the weekly quizzes and maintain an overall course grade of "A" if they received 100% of the possible points in the laboratory section of the course.

As Clayton and Madsen suggested, another reason for the lack of a difference in first quiz performance between groups may have been a result of the many procedural similarities between the remedial and

Figure 1: Median percent correct on the first weekly unit quiz for the remedial and no-remedial groups.


100
90
80
70
60
50
40
30
20
10
0
(MEDIAN)
PERCENT CORRECT
Rem
No-Rem
1
2
3
4
5
6
7
8
9
10
11
12
QUIZZES

no-remedial conditions. In both conditions, students were exposed to weekly quizzing, small units of study material, objectives, and well-sequenced materials--all components of a well-designed system of instruction (Michael, 1974).

Second Quiz Performance

Figure 2 shows the median percent correct on the second quiz contrasted with the median percent correct on the first quiz for the remedial group. Students improved their scores by studying a reported average of one hour and ten minutes for the second quiz. Some of the improvement on the second quiz may be attributable to better focused studying and the effects of practice. The median quiz score increased significantly from 89.3% correct on the first quiz to 95% correct on the second quiz, t (10) = 2.49, $p < .05$. These data support the findings of Bostow and Blumenfeld (1972), Bostow and O'Connor (1973), Clayton and Madsen (1974), Sundberg (1977), and Trainor (1977).

Figure 3 shows the median percent correct for only those students who remediated. Students significantly improved their scores by remediating. The median quiz scores increased from 70.6% on the first quiz to 87.13% on the second quiz, t (10) = 3.39, $p < .003$.

Given the opportunity, 86% of the students needing to, remediated, apparently avoiding the minimal credit received for less than "A" performance on the first quiz.

Achievement Test

The results of the achievement test show no difference in performance

Figure 2: Median percent correct on the first and second weekly unit quiz for the remedial group.

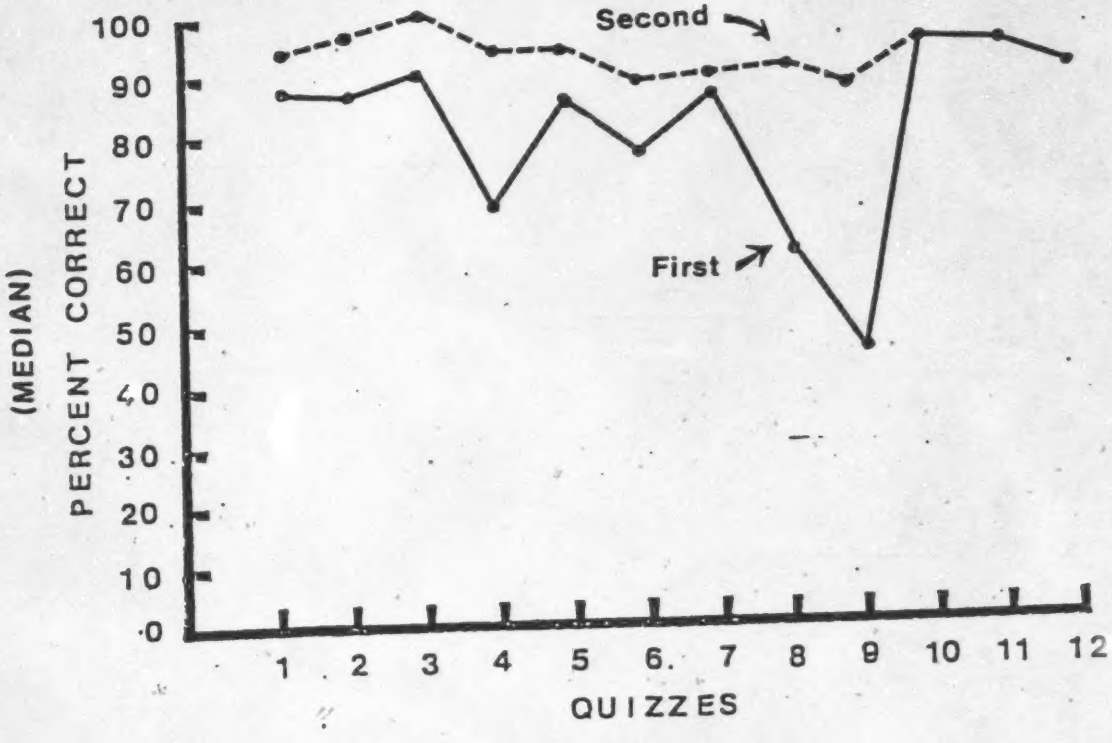
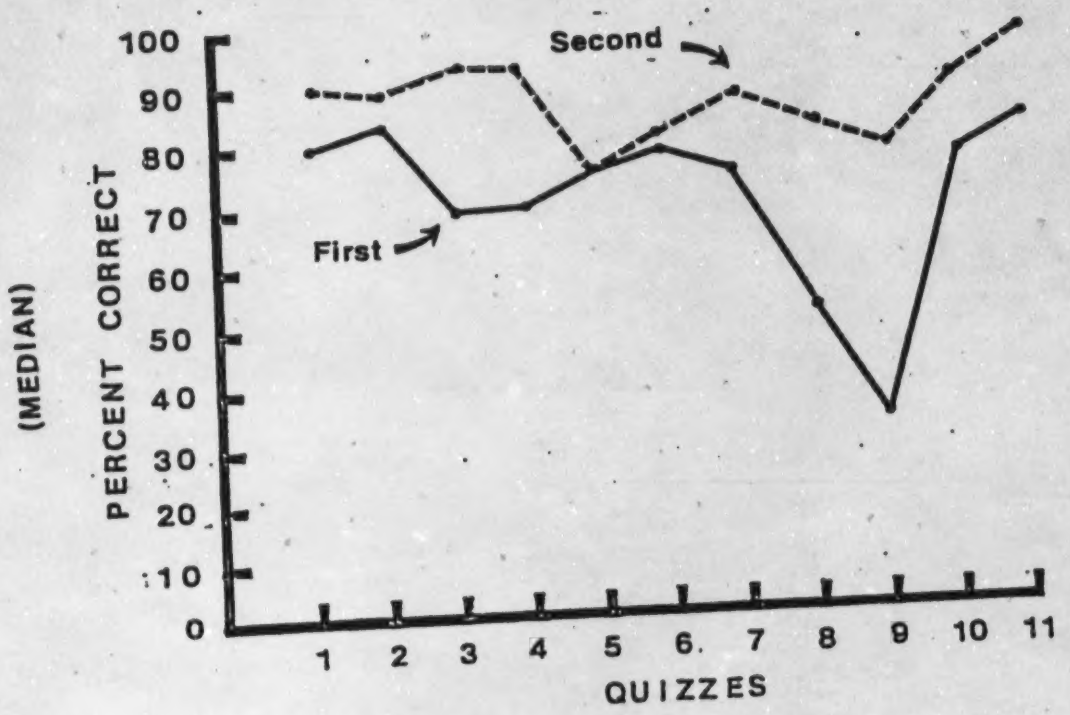

Second
First
100
90
80
70
60
50
40
30
20
10
0
(MEDIAN)
PERCENT CORRECT
1
2
3
4
5
6.
7
8
9
10
11
12
QUIZZES

Figure 3: Median percent correct on the first and second quiz for only those who remediated.


(MEDIAN)
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
Second
First
1
2
3
4
5
6
7
8
9
10
11
QUIZZES

between the remedial group and the no-remedial group. The median percent correct for the remedial group was 66, compared with 69 for the no-remedial group.

Although the students in the remedial group earned quiz scores higher than the students in the no-remedial group, a difference failed to appear on the achievement test. It may be that the sample of items tested is not directly related to what the students learned. It seems plausible to suggest the remedial group would score higher than the no-remedial group on a test consisting of every objective covered in the course. The students in the remedial group are exposed to many more test items and consequently have had more practice taking tests over the material.

Test Over Remediation Procedure

Students scored an average of 87% correct on the test given over the procedures for remediation. One question which was missed consistently asked the students to distinguish whether remediation of scores less than 90% on the initial quiz was required. However, even if students were to answer the question wrong and suggest that remediation was required, they would not have been at a disadvantage with respect to those students who answered the question correctly because in either case students would have remediated.

Staff Time

The remediation procedure required 5.3 more staff hours than the

no-remedial procedure. The overall time costs for the remediation and no-remediation procedures each week were 18.8 hours and 13.5 hours respectively.

Cost Analysis

A comparison of results and costs for a remedial quizzing procedure relative to a no-remedial procedure shows the no-remedial procedure to be more cost-efficient than the remedial procedure. This analysis is constrained by the limited scope of the outcome measures (Appendix A).

Table I shows that the remedial procedure produced 5% higher quiz scores; the "unit costs" associated with those scores are 22% higher than the "unit costs" for the scores of the no-remedial group. The large difference in "unit costs" required to produce a 5% increase in quiz scores is not cost-efficient; in other words, the no-remedial procedure, with its lower "unit costs" is more cost-efficient.

Table II shows that the remedial procedure produced slightly lower achievement test scores, of course, at a higher "unit cost" than the no-remedial procedure; therefore the remedial procedure is again less cost-efficient. The practical significance of the remedial procedure can be best represented by the number of additional objectives the students in the remedial group learned as a result of participating in remediation. Results indicated an approximate 5% relative increase in the remedial group's mean quiz scores. On a 30 point quiz, this translates into about one additional objective learned for every three

Table I: Cost-efficiency estimate.

TABLE I

COST-EFFICIENCY ESTIMATE

	GROUP		% DIFFERENCE REM VALUE - NO-REM VALUE / NO-REM VALUE
	REMEDIATION	NO-REMEDIATION	REM RELATIVE TO NO REM
1. TOTAL STAFF COST PER SEMESTER	$860.16	$616.00	+40%
2. PERCENT CORRECT ON DV	89	85	+ 5%
3. "UNIT COST" = $ / DV = (1) / (2)	$9.66	$7.90	+22%
4. COST-EFFICIENCY INDEX = DV / $ = (2) / (1)	.1	.13	-.23%

DV = Mean Quiz Score
REM = Remediation
NO REM = No Remediation

Table II: Cost-efficiency estimate.

TABLE II

COST-EFFICIENCY ESTIMATE

	GROUP		% DIFFERENCE REM VALUE - NO-REM VALUE / NO-REM VALUE
	REMEDIATION	NO-REMEDIATION	REM RELATIVE TO NO REM
1. TOTAL STAFF COST PER SEMESTER	$860.16	$616.00	+40%
2. PERCENT CORRECT ON DV	66	69	- 4%
3. "UNIT COST" = $ / DV = (1) / (2)	$13.03	$8.93	+46%
4. COST-EFFICIENCY INDEX = DV / $ = (2) / (1)	.08	.11	-27%

DV = Achievement Test (Mean Score)
REM = Remediation
NO REM = No Remediation

units, a trivial difference.

Reliability

Interobserver agreement for the weekly quizzes and the final examination grading averaged 90% over the semester with a range of 75% to 100%.

EXPERIMENT 2

The remediation procedure in Experiment 1 failed to produce significant improvement in the students' terminal performance; therefore, I designed and evaluated another procedure attempting to produce improved performance on the achievement test.

The deferred-point-scale in Experiment 1 provided different consequences for scores above and below a specified criterion, which caused a methodological difficulty. The purpose of the deferred-point-scale was to encourage students to remediate when they scored below 90% correct on the first quiz. Ultimately the combination of the opportunity to remediate and the deferred-point contingency seemed to constitute a confounding of the attempt to evaluate the remediation procedure. That is, the two groups differed along two dimensions, both the opportunity to remediate and the deferred-point-scale; either difference could have had an effect on the dependent variables.

The confounding was eliminated by removing the deferred-point-scale and giving both groups their percent correct as their grade for the quiz. In Experiment 1 it was possible for students in the remedial and no-remedial groups to receive different amounts of course credit for identical performances.

As stated, the remedial procedure for Experiment 2 required a minimum percent correct on the first quiz to qualify for remediation, motivating students to study for the first quiz without manipulating course credit. A 70% minimum performance criterion was established after

reviewing the past semester's quiz scores and finding that approximately 85% of the first quiz scores were above 70%. With the minimum performance criterion for the opportunity to remediate, students should study before the first quiz rather than procrastinate and only study just before the remedial quiz.

Even with the strong motivational system in Experiment 1, some students failed to achieve an "A" on the first quiz. It is possible that, either they did not study at all, or that the individual educational histories of some students are such that they require more practice or different arrangements of instructional stimuli in order to respond appropriately to questions concerning topics in the reading. In an attempt to provide contingencies which bring the students into more structured contact with the material and ensure studying, the remediation procedure in Experiment 2 specified that students had to answer, in written form, a list of study questions at 90% correct and attend a remedial lecture/discussion period before being permitted to take the remedial quiz.

The remedial study questions required students to respond to novel examples of the concepts in the reading and also to develop novel examples of those concepts. Some of the text books provided study questions that only required students to recall content from the reading, not encouraging students to come under the control of a wide range of the stimulus features within each concept.

The remedial study questions structured students' studying somewhat, possibly bringing them in contact with the reading in such a way as to increase their second quiz scores. The students who did not

study before the first quiz would now presumably have to study at least those topic areas covered by the remedial study questions, before retesting. Additionally, the students who studied but did not reach an "A" would now have the opportunity to have their study guided by the remedial study questions and would have practice responding in written form to questions over the reading. As indicated by Johnson and Ruskin (1977), requiring students to answer study questions before testing increases students' performance on written and oral tests.

Concomitant with the procedural changes, two major changes in research methodology occurred. The course ATA recorded another form of grading reliability based on samples of grading taken without the graders knowledge, adding one more check on grading accuracy. Also, based on Sultzer-Azaroff's 1976 procedure, I constructed individualized probe tests containing items missed on each first quiz. The method describes the differences in procedures between Experiments 1 and 2.

Method

Subjects

Before asking for volunteers in Applied Behavior Analysis Lab I, I announced that the purpose of the study was to develop an effective remediation procedure. Further, I elaborated on the failure of several previous attempts to demonstrate the effectiveness of similar remedial procedures in the course and that all volunteers would be randomly assigned to either the standard no-remedial quizzing procedure or to the experimental remedial quizzing procedure. Out of a class of 58,

a total of 36 students signed informed consent forms and volunteered to participate. On the basis of grade point averages, I matched each student with another student with similar grade point average. Then, one member of each pair was randomly assigned to each procedure. The median grade point average of the no-remedial group was 3.0 and the median grade point average of the remedial group was 3.03. On a pretest over the first five units of the course both groups scored a median number correct of six out of a possible 20 questions.

Remedial Group

Students in the remedial group had the option of participating in the weekly remediation activities. All students took the first quiz at the same time. Students had to score 70% or above to qualify for the remedial activities. The weekly remedial activities included: (1) completing study questions over the unit at a minimum of 90% before the remedial lecture/discussion period, (2) attending the remedial lecture/discussion period, and (3) taking the second form of the quiz. Students failing to satisfy the criteria were not permitted to remediate. Students who scored 70% or more but did not remediate kept their first quiz score. While those who remediated kept the best of the two scores.

No-Remedial Group

The students in the no-remedial group did not have the option of participating in the weekly remedial activities but instead retained their percent correct on the first quiz as their score for the unit.

All other quizzing procedures were the same as those for the remedial group.

Dependent Variable

Achievement test students took the achievement test in sections. During the first week of the study, students took the first half of the test which covered the first four weeks of material; then, during the fourth week of the study, students took the entire achievement test. Finally, on the day of the last weekly quiz, students took the last half of the test which covered the remaining five units of material.

Individualized probe test. Students took an individualized probe test during the last week of class along with the last section of the achievement test. The purpose of giving the students the probe test was to determine whether students in the remedial group were more likely to learn the quiz questions they originally missed on the first quiz than the students in the no-remedial group. I constructed an individualized probe test for each student in both groups. All of the students who had missed ten or more questions received a probe test containing ten randomly selected questions from the pool of questions they missed on all of the first quizzes, all of the students who missed ten or fewer questions received a probe test containing all of the questions they missed on the first quizzes.

The TAs graded the probe tests following the same grading-reliability procedure used for quizzes and the final examination.

Students received five points for completing the probe and achievement tests--these points replaced points on the last weekly quiz. No

points were contingent upon the level of performance on these tests.

Reliability of the dependent variables. The reliability procedures for Experiment 2 contained a component not present in Experiment 1; all other reliability procedures were the same as in Experiment 1. In addition to the reliability procedures previously specified, the ATA attached carbon paper to a random sample of approximately 25% of the week's quizzes and graded the carbon copies on four occasions throughout the semester.

Results and Discussion

The results of Experiment 2 indicate that students significantly improved their weekly quiz scores when they remediated. And, students who remediated were more likely to answer correctly questions they had previously answered incorrectly on the first quiz.

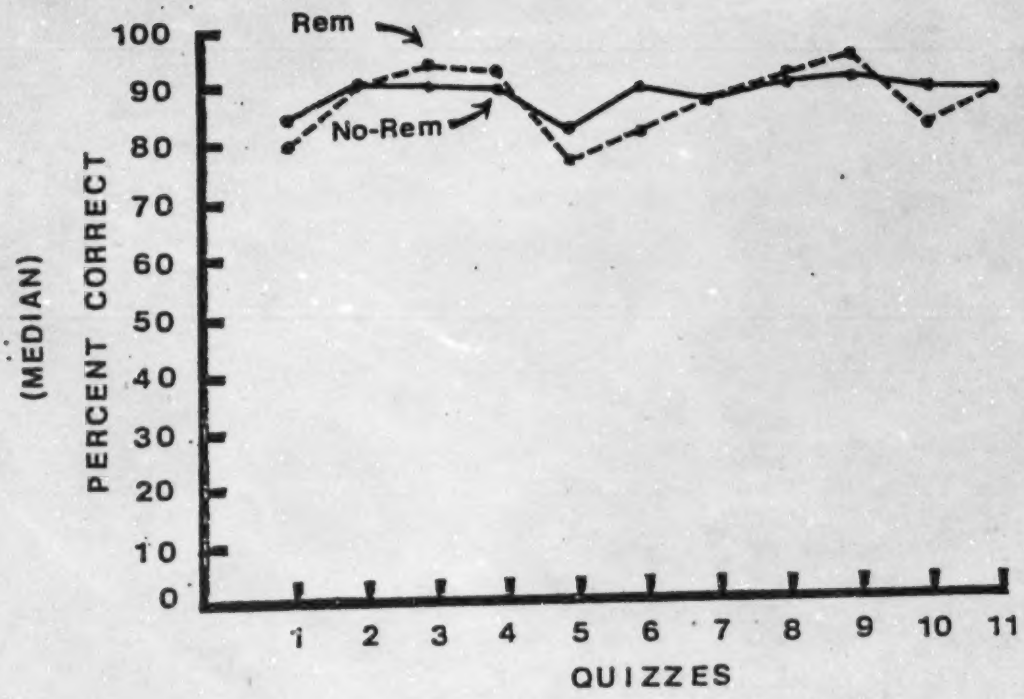
First Quiz Performance

Figure 4 shows the median percent correct on the first quiz for the remedial and no-remedial groups. The data indicate, as in Experiment 1, no systematic differences across the semester between groups, as the median percent correct per group was the same, 90%.

As in Experiment 1, students who participated in the remedial group reported they studied more, on the average, than students in the no-remedial group--3 hours and 20 minutes compared to 3 hours and 10 minutes for each initial quiz.

The first quiz performance approximated the criterion for a course grade of "A", as in Experiment 1.

Figure 4: Median percent correct on the first quiz for the remedial and no-remedial groups.


Rem
No-Rem
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
(MEDIAN)
1
2
3
4
5
6
7
8
9
10
11
QUIZZES

Second Quiz Performance

Figure 5 shows the remedial group's quiz scores improved insignificantly from a median of 90% correct on the first quiz to a median of 93% correct for the unit.

Students in the remedial group remediated only 39% of the time the opportunity was available marking a sharp decrease from the 86% in Experiment 1.

Figure 6 clearly shows that those students who remediated significantly improved their scores for the unit from a median of 81.3% correct to 92.8% correct, t (8) = 16.8, $p < .001$. However, as in Experiment 1, this improvement in weekly quiz scores did not lead to a systematic difference for the remedial group on the achievement test.

Students reported studying an average of 45 minutes in preparation for the second quiz. Note that in Experiment 1, students reported that they studied 1 hour and 10 minutes--it could be that students in Experiment 2 studied more efficiently with the aid of the remedial study questions or that the consequences for not studying were less severe. In that in Experiment 1 students who scored below 67% on the first quiz received no credit for the unit without remediating, in Experiment 2, students who scored 67% would not be permitted to remediate and were required to keep their 67% as the score for the unit.

Achievement Test

Table III shows the mean percent correct on the pre- and post-achievement tests. There was no significant difference between the two

Figure 5: Median percent correct on the first and second quizzes for the remedial group.

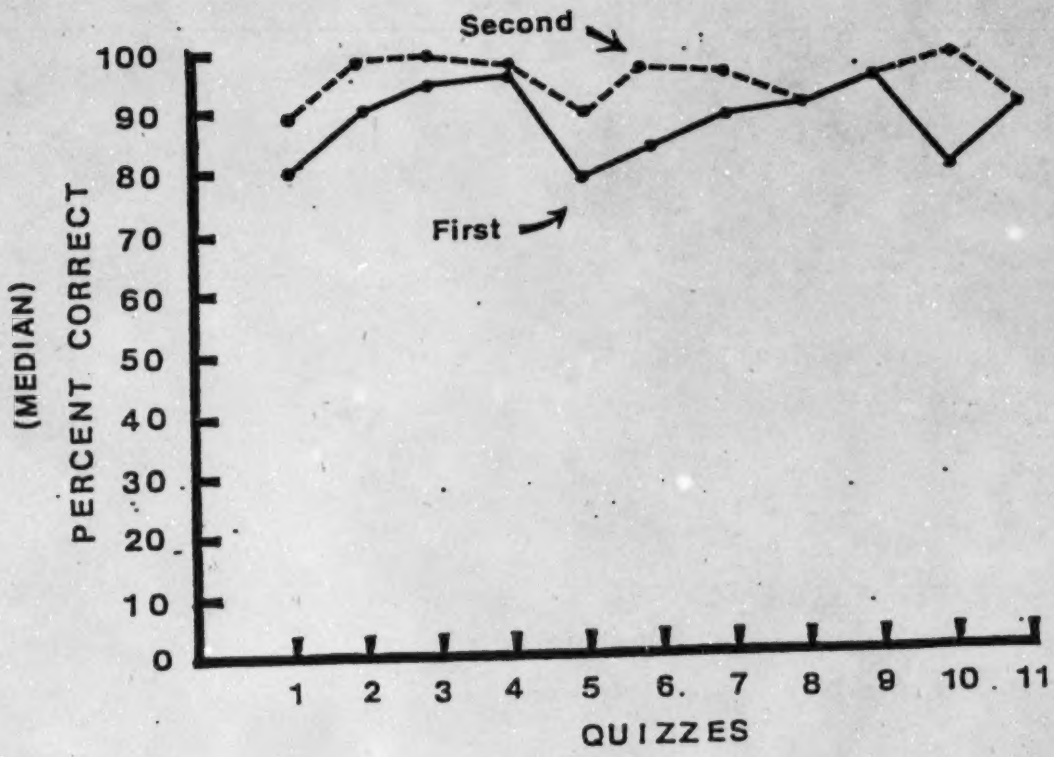

Second
First
100
90
80
70
60
50
40
30
20
10
0
(MEDIAN)
PERCENT CORRECT
1
2
3
4
5
6.
7
8
9
10
11
QUIZZES

Figure 6: Median percent correct on the first and second quizzes for only those who remediated.

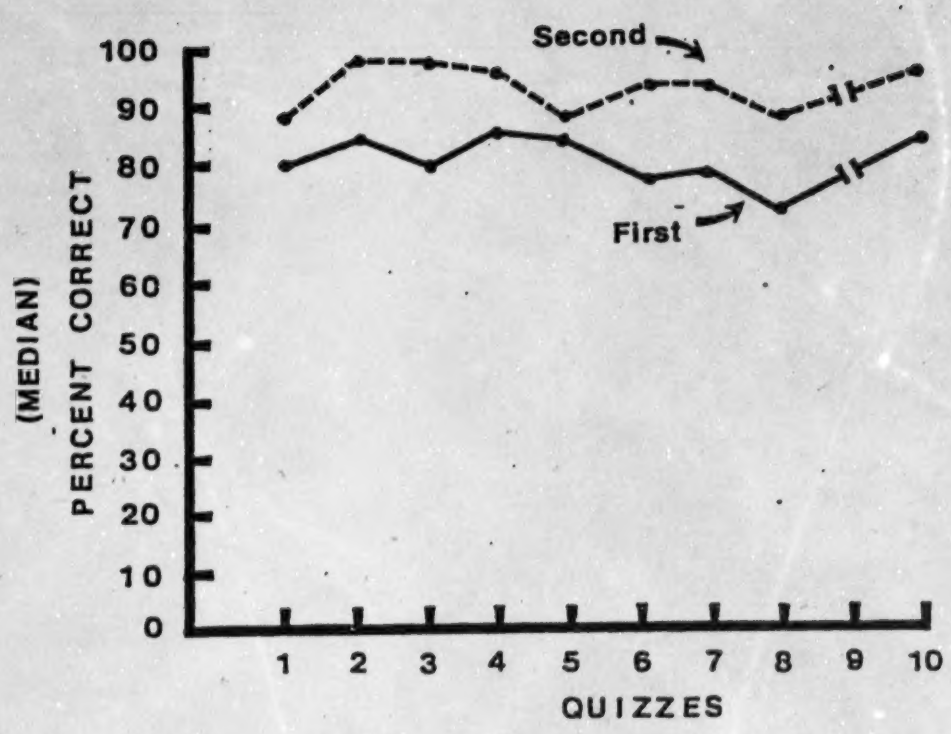

Second
First
100
90
80
70
60
50
40
30
20
10
0
(MEDIAN) PERCENT CORRECT
1
2
3
4
5
6
7
8
9
10
QUIZZES

Table III: Mean percent correct on the achievement test.

TABLE III

MEAN PERCENT CORRECT ON THE ACHIEVEMENT TEST

	Pretest 1	Posttest 1	Pretest 2	Posttest 2
Remedial Group	38.2	67.9	48.9	74
No-Remedial Group	38.8	58.5	49.3	74.4

groups or matched pairs on the achievement test as a function of the remediation procedure.

Individualized Probe Test

Figure 7 shows the median percent correct for the remedial group was significantly greater than the median percent correct for the no-remedial group, t (12) = 17, $p < .05$. The remedial and no-remedial groups scored a median of 71% correct and 33% correct respectively. In 9 out of 13 pairs, the student in the remedial group scored higher. The probe test is biased toward good results for the remedial group as it tests items the students reviewed for the remedial quiz. Nevertheless, the difference is large and indicates that students do learn more of the test items they missed on the first quiz if they participate in remedial activities and retest in a lecture course.

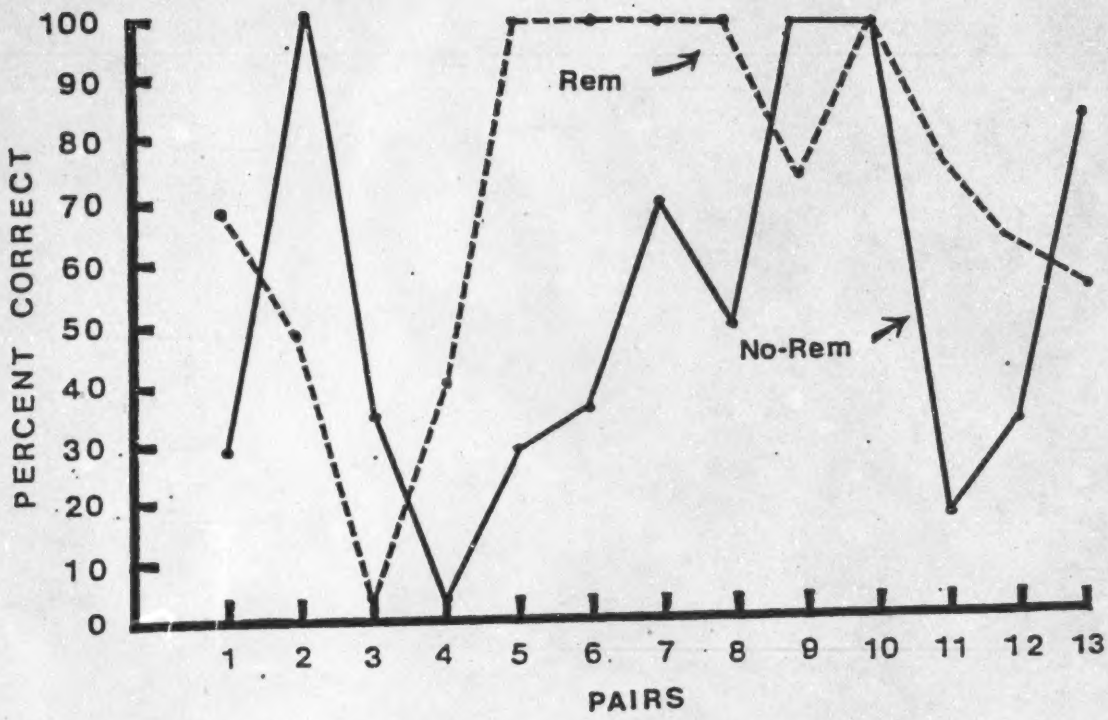
Staff Time

The remediation procedure required 4 hours and 20 minutes more staff time a week than the no-remedial procedure. This additional staff time was spent on remedial quiz grading, proctoring, and conducting the remedial lecture/discussion period. Administration of the first quiz only, cost 11.6 staff hours; whereas, the total administrative costs for the no-remediation procedure was 15.8 staff hours.

Course Evaluation

When asked the question, "How do you feel about the teaching procedures in this course?" on the department course evaluation, 74% of

Figure 7: Median percent correct on the individualized probe tests for the remedial and no-remedial groups.


100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
Rem
No-Rem
1
2
3
4
5
6
7
8
9
10
11
12
13
PAIRS

the students in the remedial group responded with the options, "I am well satisfied with these teaching procedures" or "The course provided me with a valuable learning experience," as compared to 37% of the students in the no-remedial group.

Cost Analysis

This cost analysis supports the results of Experiment 1 showing the no-remedial procedure to be more cost-efficient in terms of achievement test and quiz performance. However, this analysis shows the remedial procedure to be more cost-efficient in terms of individual probe test performance. Again, as in Experiment 1, this estimation is constrained by the scope of the outcome measures (Appendix A). (See Tables IV and V.)

Table VI shows the remedial procedure generated 115% higher individualized, probe-test scores, relative to the no-remedial procedure, while increasing "unit costs" 64% and, thereby producing 43% better cost-efficiency. These findings conflict with the findings for cost-efficiency on all other measures.

This difference in the results of the analysis on the individualized probe tests illustrates the primary function of the remedial procedure. It is evident that the remedial procedure makes it more likely that the students learn previously unlearned material, and it does so in a more cost-efficient manner than the no-remedial procedure.

Consideration should be given to the practical implications of the cost analysis on the individualized probe test data. This analysis shows 115% higher test scores for the remedial group, which represents

Table IV: Cost-efficiency estimate.

TABLE IV

COST-EFFICIENCY ESTIMATE

	GROUP		% DIFFERENCE REM VALUE - NO-REM VALUE / NO-REM VALUE
	REMEDIATION	NO-REMEDIATION	REM RELATIVE TO NO REM
1. TOTAL STAFF COST PER SEMESTER	$692.80	$504.64	+37%
2. PERCENT CORRECT ON DV	90	87	+ 3%
3. "UNIT COST" = $ / DV = (1) / (2)	$7.70	$5.80	+33%
4. COST-EFFICIENCY INDEX = DV / $ + (2) / (1)	.12	.17	-.29%

DV = Mean Quiz Score
REM = Remediation
NO REM = No Remediation

Table V: Cost-efficiency estimate.

TABLE V

COST-EFFICIENCY ESTIMATE

	GROUP		% DIFFERENCE REM VALUE - NO REM VALUE / NO REM VALUE
	REMEDIATION	NO-REMEDIATION	REM RELATIVE TO NO REM
1. TOTAL STAFF COST PER SEMESTER	$692.80	$505.64	+37%
2. PERCENT CORRECT ON DV	61	64	-.05%
3. "UNIT COST" = $ / DV = (1) / (2)	$11.36	$7.89	+44%
4. COST-EFFICIENCY INDEX = DV / $ = (2) / (1)	.09	.13	-30%

DV = Achievement Test (mean score)
REM = Remediation
NO REM = No Remediation

Table VI: Cost-efficiency estimate.

TABLE VI

COST-EFFICIENCY ESTIMATE

	GROUP		% DIFFERENCE REM VALUE - NO-REM VALUE / NO-REM VALUE
	REMEDIATION	NO-REMEDIATION	REM RELATIVE TO NO REM
1. TOTAL STAFF COST PER SEMESTER	$692.80	$504.64	+37%
2. PERCENT CORRECT ON DV	71	33	+115%
3. "UNIT COST" = $ / DV = (1) / (2)	$9.75	$15.29	-36%
4. COST-EFFICIENCY INDEX = DV / $ = (2) / (1)	.1	.07	+43%

DV = Probe Test (mean score)
REM = Remediation
NO REM = No Remediation

actual mean scores of 71% and 33% for the remedial and no-remedial groups. These means scores were derived from a maximum of ten quiz questions per student in each group. So the 115% higher scores indicate that the average student in the remedial group may have learned as few as four or five more quiz objectives over the entire semester at a cost of $188 for the semester.

Test on Remediation Procedure

Students scored a median of 89% correct on the test covering remediation procedure guidelines.

Reliability

Interobserver agreement for the weekly quizzes and the final examination averaged 93% over the semester with a range of 84% to 100%.

Interobserver agreement for the covert reliability averaged 94% agreement with a range of 85% to 98% for the checks.

CONCLUSIONS

Students tended to score at or near the "A" level on the first quizzes in the present course, with its frequent quizzing, small units of study, clearly written text material, and unit study questions. So it seems plausible that raising the criterion for an "A" would increase the students' performances.

Their performance can improve further with the use of the proper remedial procedures. This was seen in Experiment 2, where students in the remedial group scored twice as well on the individualized probe test as the students in the no-remedial group. In view of these findings, it seems reasonable to motivate students to remediate those test items they missed on the first quiz, if the goal of the course is to teach a specific and well-defined set of concepts. In the long run, it may be to the advantage of the students to remediate those concepts they failed to master, even though it may be more cost-efficient, over the semester, not to include a remediation procedure in the course.

We succeeded more in motivating students to remediate when minimal course credit was assigned for less than "A" performance on the first quiz than when meeting a minimum performance criterion was necessary to qualify for remediation. So, it seems that instructors can motivate students to remediate scores of less than 90% correct by restructuring the point system so that retaining the "A" is not possible without scoring 90% or better in all course requirements.

REFERENCES

Blair, P. A comparative cost analysis of PSI and conventional teaching. A paper presented at the Fourth National Conference on Behavior Research and Technology in Higher Education, 1977.

Bostow, D. E. and Blumenfeld, G. J. The effect of two-test retest procedures on the classroom performance of undergraduate college students. In George Semb (Ed.), Behavior Analysis and Education. Lawrence, Kansas: Support and Development Center for Follow Through, Department of Human Development, University of Kansas, 1972.

Bostow, D. E. and O'Connor, R. S. A comparison of two college classroom testing procedures: required remediation versus no remediation. Journal of Applied Behavior Analysis, 1973, 6, 599-607.

Clayton, T. G. and Madsen, C. H. Jr. A comparison of three methods of teaching child psychology using a progressive multiple baseline achievement test. Research and Technology in College Teaching, 1974, 2, 205-213.

DuNann, D. H. and Fernald, P. S. An experimental comparison of a contingency managed course with a large lecture method course. Journal of Applied Behavior Analysis, 1976, 373-374.

Geis, G. L., Stebbings, W. C. and Lundin, R. W. Reflex and Operant Conditioning. New York, New York: Appleton-Century-Crofts, 1965.

Hayman, R. W. and Levin, H. Economic analysis and historical summary of educational technology cost. In Productivity and Efficiency in Education, Washington, D. C.: Federal Council on Science and Technology, 1973.

Hursh, D. E. Personalized systems of instruction: what do the data indicate? Journal of Personalized Instruction, 1976, 1, (2), 91-105.

Johnson, J. M. and O'Neill, G. The analysis of performance criteria defining course grades as a determinant of college academic performance. Journal of Applied Behavior Analysis, 1973, 6, 261-268.

Johnson, K. R. and Ruskin, R. S. Behavioral Instruction: An Evaluative Review. American Psychological Association, Washington, D. C., 1977.

Katzenberg, A. C. How to Draw Graphs. Kalamazoo, Michigan: Behavior-delia, Inc., 1975.

Malott, R. W., General, D. A. and Snapper, V. B. Issues in the Analysis of Behavior. Kalamazoo, Michigan: Behaviordelia, Inc., 1973.

Malott, R. W. and Svinicki, J. G. Contingency management in an introductory course for one thousand students. The Psychological Record, 1969, 19, 545-556.

McSween, T. A point system for maintaining accurate grading in a PSI course. Unpublished manuscript, Western Michigan University, 1976.

Michael, J. Behavior Modification: Applications to Education. New York, New York: Academic Press, Inc., 1974.

Miller, L. K. and Weaver, F. H. A multiple baseline achievement test. In George Semb (Ed.), Behavior Analysis and Education. Lawrence, Kansas: Support and Development Center for Follow Through, Department of Human Development, University of Kansas, 1972.

Reese, D. G. and Woolfenden, R. M. Behavioral Analysis of Everyday Life: A Program for the Generalization of Behavioral Concepts. Kalamazoo, Michigan: Behaviordelia, Inc., 1973.

Semb, G. The effects of mastery criteria and assignment length on college student test performance. Journal of Applied Behavior Analysis, 1974, 7, 61-69.

Sultzer-Azaroff, B., Johnson, Dean, Freyman, An experimental analysis of proctor quiz-scoring in PSI courses. Journal of Personalized Systems of Instruction, 1977, 1, (2), 143-149.

Sundberg, C. M. An examination of the different effects of retesting and mastery criterion on student performance in a PSI course. Unpublished manuscript, Western Michigan University, 1977.

Trainer, T. N. The effect of an individualized remediation program on comprehensive final examination performance. Unpublished manuscript, Western Michigan University, 1977.

APPENDIX A

Cost Analysis

The value of educational research as a tool for administrative decision-making can be improved by including cost factors in the analysis. It is clear that administrators respond to more than student achievement when assessing the value or effectiveness of educational plans--it is appropriate to consider the resource expended for a given effect on performance. The most convenient representation of expended resources is the number of dollars spent.

The standard economic measuring tools are benefit/cost comparisons, cost-effectiveness computations, cost-efficiency estimates, and cost-utility calculations (Hayman and Levin, 1973).

Benefit/cost comparisons require educational benefits to be measured directly in money earned or money saved by the project or procedure, which is difficult to predict accurately. Also, time constraints must be included in the analysis. So, benefit/cost comparisons are not reasonable in the present case, because figuring the students' financial benefit is not currently feasible.

A cost-effectiveness analysis may be used in cases where the goal is maximizing the effectiveness of a given budget or minimizing cost while maintaining performance. In a cost-effectiveness one unambiguously defined outcome must be specified as a goal, in order to judge which alternative is most desirable. This specified outcome must be the same for each procedure. It is not clear that any one of the outcomes

(dependent variables) is sufficient to serve as a goal for the present research. For lack of one defined outcome we move to cost-efficiency as the appropriate cost analysis for this research.

A cost-efficiency estimation requires only that the outcomes be quantifiable and allows procedures to be analyzed over a given period of time. Cost-efficiencies may be compared among several projects along several outcomes. The outcome (in the present study, the various dependent variables) divided by input (staffing costs in dollars) yields several indeces or cost-efficiency estimations.

Of course, the power of the cost-efficiency estimation is constrained to the extent that the outcome measures are not relevant to the overall goals of the procedure. Consideration should be given to the scale and the time span of the procedure, when using cost-efficiency, because the cost-efficiency may reverse when the breadth of the procedure changes. Cost-efficiency seems to be applicable here; and, therefore, tables of results appear in each experiment.

Another cost measure, cost-utility, weighs subjective judgement in its analysis and is most useful when the methods above fail to show a difference between alternatives (Blair, 1977).

APPENDIX B

Recommendations for Psychology 351

1. Do not include remediation in course plan.

2. Raise criteria for "A" by including larger units of reading material or including more units. (Students are not spending enough time with course work to justify the amount of credit they receive.)

3. Require students who perform poorly (below 90%) for two consecutive units to complete extra course work to qualify for the next quiz, thus motivating study before the quiz.

APPENDIX C

Recommendations for Future Research

1. Establish high criteria for "A" before starting to make certain students actually need to remediate.

2. Equate difficulty of quizzes.

3. Develop a more comprehensive achievement test which directly samples a larger portion of the course objectives.

4. Match students by pretest or performance in similar courses.

5. Use computer storage and data analysis.

6. Require students who remediate to keep their remedial score to motivate studying.

7. Evaluate students' study skills, reading speed, grammatical ability, self-management by requiring a study log and an English test.

8. Before attempting to study or design a remediation procedure, check to see if the course is either highly sequential, or one in which it is necessary for all students to reach a high level of competency. If not, remediation may not be essential and research on other components of the course may be more beneficial.